

# DOCUMENT RESUME

ED 075 489

TM 002 568

TITLE Massachusetts Fourth Grade Testing Program 1971.  
 INSTITUTION Massachusetts State Dept. of Education, Boston.  
 REPORT NO Test-Bull-1  
 PUB DATE Apr 71  
 NOTE 4p.

EDRS PRICE MF-\$0.65 HC-\$3.29  
 DESCRIPTORS \*Achievement Tests; Aptitude Tests; Basic Skills;  
 \*Grade 4; \*Measurement Instruments; State Programs;  
 Statistical Analysis; \*Testing Programs; \*Test  
 Results  
 IDENTIFIERS \*Massachusetts

## ABSTRACT

The testing of every fourth-grade classroom in Massachusetts was carried out in an effort to answer the following questions: (1) What are the levels of mastery of basic skills in Massachusetts fourth grades? Are there differences in achievement between skills?; (2) What educational needs can be inferred for Massachusetts' students, based on basic skills testing?; (3) Do testing data reveal the influence of Federal programs?; (4) Does the product of education vary according to available resources--financial outlay, professional support, materials?; and (5) Are there regional variations in abilities and achievement? Aptitude and achievement data were obtained for 324 school systems, 1488 schools, and 85,382 fourth-grade children. The test instruments used were the Comprehensive Tests of Basic Skills and the Short Form Test of Academic Aptitude published by CTB/McGraw-Hill. Three different reports of the test data were supplied to all school systems. The test data showed that the state as a whole exceeded the national norms; the mean "obtained" scores were significantly higher than the "anticipated" scores in all areas measured by the tests of basic skills. Highest scores were in reading comprehension and the lowest in arithmetic. From the test results, it was concluded that Massachusetts fourth graders are slightly higher than the national norms in all areas measured. Correlations between the subtest total mean scores by school are statistically significant and very high. Schools that did well on one subtest generally did well on all subtests. A survey of school superintendents showed that 98% used the test data. (DB)

April, 1971

U.S. DEPARTMENT OF HEALTH,  
EDUCATION & WELFARE  
OFFICE OF EDUCATION  
THIS DOCUMENT HAS BEEN REPRO-  
DUCED EXACTLY AS RECEIVED FROM  
THE PERSON OR ORGANIZATION ORIG-  
INATING IT. POINTS OF VIEW OR OPIN-  
IONS STATED DO NOT NECESSARILY  
REPRESENT OFFICIAL OFFICE OF EDU-  
CATION POSITION OR POLICY

## MASSACHUSETTS FOURTH GRADE TESTING PROGRAM 1971

### Test Bulletin #1

#### Background information

In August 1970, Commissioner of Education Neil Sullivan announced plans to test all Massachusetts fourth graders during January 1971. The department had been moving toward this since early 1970, when the Bureau of Curriculum Innovation assembled a committee to investigate methods of Needs Assessment, as related to the joint evaluative concerns of both Title I and Title III. In June 1970, another group was organized to rewrite state NDEA-VA into ESEA Title III.

Areas explored by this group included procedures for determining educational or instructional objectives, including Massachusetts' involvement with the Instructional Objectives Exchange at UCLA; adopting a standardized testing program; as well as alternatives to a standardized testing program. These latter alternatives involved local selection of instructional objectives either from a prepared list of objectives or from local specification of objectives teamed with some form of evaluation.

Techniques varied from asking a test publisher such as Educational Testing Service to prepare a series of comparable instruments from their item bank by using item sampling techniques, to working with Project Comprehensive Achievement Monitoring to tailor assessment to the individual school system. Analyzing the results of each system's testing program was also explored. There appears to be at present no direct way to compare the results of the several testing programs. The standardization and norms vary for each of the testing programs. The time of testing in Massachusetts's schools varies as well. Some systems conduct limited testing while others are extensive. There is neither overlap nor comparability among local testing programs.

#### Description of the Program

The key questions to be answered by the testing program were:

- What are the levels of mastery of basic skills in Massachusetts fourth grades? Are there differences in achievement between skills?
- What educational needs can be inferred for Massachusetts' students, based on basic skills testing?
- Do testing data reveal the influence of Federal programs?
- Does the product of education vary according to available resources - financial outlay, professional support, materials?
- Are there regional variations in abilities and achievement?

One of the greatest problems with a program of this magnitude, involving every fourth grade classroom in Massachusetts, was the communication barrier between the State Department and individual teachers. In order to inform all

Concerned, meetings were planned with superintendents for their regular round-table meetings in September and October. Pretest workshops for system test coordinators were scheduled by geographic area during November to present information about the instruments, the interpretation of test scores, and the distribution and collection of materials. The Assistant Commissioner for Research and Development taped a 15 minute television program to discuss the testing with fourth grade teachers. Channel 2 (Boston) presented this tape on three consecutive afternoons for the eastern part of the state and on closed circuit television for the western half.

### Instruments and Reports

The instruments selected for this program were the Comprehensive Tests of Basic Skills (CTBS) and the Short Form Test of Academic Aptitude (SFTAA), published by CTB/McGraw-Hill. CTBS provided information on levels of mastery for both learning content and process areas. When used in relation with SFTAA, CTBS provided information on actual or "obtained" achievement compared with potential or "anticipated" achievement for individuals, classes, schools and systems. These "anticipated" achievement scores were computed for each student using multiple regression formulas which utilize certain predictors - age, grade in school, sex and raw scores for each of the subtests of the California Short-Form Test of Academic Aptitude.

The Commonwealth supplied three different reports of test data for all school systems:

- The "Administrator's Summary of Test Data Mean Values" which reported subtest and total subtest mean, "obtained" and "anticipated" scores for each school as well as for the district as a whole.
- The "Combination Class Record" which reported "obtained" and "anticipated" achievement scores for each class member and summary data for each class.
- The "Summary Report of the Right Response Record and Item Analysis" which presented group item mastery data for both content and process dimensions.

Testing was conducted during the second week in January. Post-testing workshops were conducted at the Department of Education Regional Centers during the last half of March. Resource materials were prepared by Research and Development for these workshops to enable local educational agencies to effectively utilize test results. At this time it was suggested that teachers focus on the difference score which compared the "obtained" and "anticipated" scores plus the national percentile rank of the "obtained" achievement score for each student. Through use of subtest scores plus tables summarizing item content and process dimensions, teachers could identify individual student instructional needs. System level evaluations relative to overall instructional goals could be made based on the Right Response Summary and the Administrator's Summary data.

### Analysis of Data

Aptitude and achievement data were obtained for 324 school systems, 1488 schools, and 85,382 children. This impressive data bank provided the basis for statistical analyses to answer questions verbalized in the initial planning stages for the program as well as those evolved as a consequence of the testing.

Qualifying statements were necessary to view the results of the testing program in the reality of the whole educational process:

- Different school systems spend different amounts of time on the basic skills.
- Measures were obtained in the basic skills; not in the content fields, art and music, or attitudes and values.
- Results of this testing cannot be considered an evaluation of teaching per se, as teaching effectiveness interacts with instructional materials and support services, administrative and supervisory leadership, parental and community support, and cultural contributions of the immediate environment.

### Statewide Conclusions

The State data showed that 85,382 fourth graders reflected a mean chronological age of 9 years, and a mean total IQ of 106. The State as a whole exceeded the national norms. The mean "obtained" scores (ESS) were significantly higher than the "anticipated" scores (AASS) in all areas measured by the tests of basic skills. Highest scores were obtained in reading comprehension and lowest arithmetic computation.

Based on CTB/McGraw-Hill analyses generalizations could be drawn regarding student performance at three different IQ levels, 113 and above, 96-104, and 87 and below. The difference in "obtained" versus "anticipated" scores was significantly higher for the below average and above average groups than for the average ability group, although all group "obtained" scores were significantly higher than "anticipated" scores.

The "obtained" total battery achievement score was compared with elementary instructional expenditures across districts. The prorated elementary cost figures were rank ordered. The top 25%, consisting of systems spending \$525 and above an instruction, were designated high cost systems. The bottom 25%, consisting of systems, spending \$425 and below, were designated low cost systems. The average expenditure figure was \$384 for the low cost systems and \$594 for the high cost systems. A T-test indicated no significant difference in achievement in basic skills between high cost and low cost systems.

Based on the "Right Response Record and Item Analysis" levels of mastery were averaged across all subtests and all districts. The means ranged from 73% to 84% of the students answering items correctly. It was concluded, therefore, that Massachusetts fourth graders are slightly higher than the national norms in all areas measured.

Correlations between the subtest total mean scores by schools are statistically significant and very high. Schools that did well on one subtest generally did well on all subtests. Consistent with other analyses of test results there was no relation between high expenditure instructional cost and high achievement in the basic skills.

### Interpretation Aids

The following analyses were distributed to superintendents:

- A stanine scale based on difference scores across all districts enabled an individual school system to view itself in relation to State performance. Difference scores were chosen as they control for different aptitude levels. The same procedure was followed to prepare school stanine scales.
- T-tests were computed to determine the significance of the difference scores for each subtest by system. This too, was prepared for each school.
- The number and percentage of over, and under and average achievers for each system within each subtest was computed.
- The mean mental age for each system was also included.

### Results of Test Use Survey

A survey was sent to all Massachusetts superintendents to ascertain the extent of use of fourth grade testing data. Ninety-eight (98%) replied that the data had been used. The following table indicates to whom data were reported and the extent of in-service use:

|                      |     |
|----------------------|-----|
| REPORTED TO          |     |
| SCHOOL COMMITTEE     | 85% |
| TEACHERS/PRINCIPAL   | 92% |
| PARENTS              | 35% |
| NEWSPAPERS           | 7%  |
| CURRICULUM PLANNING  | 52% |
| IN-SERVICE WORKSHOPS | 49% |

Other methods of use included: planning enrichment for middle groups, student conferences, grouping for instruction, comparison with other testing results, and counselor review of underachievement.

### Classroom Visitation

Members of the Department of Education visited twenty-nine schools across Massachusetts which exhibited high positive difference scores in reading, language, and arithmetic. The purpose of the visitations was to identify factors which may be associated with high achievement in basic skills. These schools represented a variety of socio-economic patterns and were located in inner-city, suburban and rural areas.

It was the conclusion of the visitation teams that instruction in these fourth grades was generally geared to the basic skills, even in the content areas. Daily routines were well established providing an economic use of learning time: children were aware of the limits within which they functioned. A high degree of respect - teacher for pupil, pupil for pupil and pupil for teacher was exhibited in most classrooms.

### Conclusion

This report describes an initial attempt to assess through testing by the Massachusetts State Department of Education. Future programs based on a heightened responsibility both to the profession and the public, will reflect measurement in a greater variety of areas.